

# On the Capacity of Noisy Computations

François Simon

Institut TELECOM ; Telecom SudParis ; CITI  
9 rue Charles Fourier, 91011 EVRY Cedex, France  
Email: francois.simon@it-sudparis.eu

**Abstract**—This paper presents an analysis of the concept of *capacity for noisy computations*, i.e. algorithms implemented by unreliable computing devices (e.g. noisy Turing Machines). The capacity of a noisy computation is defined and justified by companion coding theorems. Under some constraints on the encoding process, capacity is the upper bound of *input rates* allowing reliable computation, i.e. decodability of noisy outputs into expected outputs. A model of noisy computation of a perfect function  $f$  thanks to an unreliable device  $F$  is given together with a model of reliable computation based on input encoding and output decoding. A coding lemma (extending the Feinstein's theorem to noisy computations), a joint source-computation coding theorem and its converse are proved. They apply if the input source, the function  $f$ , the noisy device  $F$  and the cascade  $f^{-1}F$  induce AMS and ergodic one-sided random processes.

## I. INTRODUCTION AND RELATED WORKS

Reliable computation with unreliable devices, or in the presence of noise, has been the subject of numerous works within the vast field of fault-tolerant computing. Computation can be made reliable using information and component/gate redundancy. Some works aim at identifying theoretical "boundaries" on the amount of necessary and/or sufficient redundancy to achieve reliability. Recent references (see for example, [1], [2], [3], [4]) continue to extend the stream opened by Von Neumann's seminal paper [5]. These works identify bounds (e.g., depth and size of circuits) and propose frameworks to design reliable computations mainly thanks to gate redundancy. These papers do not address the question of boundaries about information redundancy. This subject, through the concept of capacity and coding theorems, has been thoroughly studied for data communication. It has not been the case for computation although the two problems (noisy computation and noisy communication) are very close: the matter is to retrieve expected values from the outputs of a random process.

The question of whether a noisy computation possesses a capacity (or equivalently whether some coding theorems for noisy computation hold and in which cases) has been raised by P. Elias in 1958 [6]. There is a practical consequence in answering positively this question. This would mean that, given a noisy implementation of an expected (e.g., Turing computable) function, it is possible to find families of efficient input codes which asymptotically allow an almost perfect computation. Efficiency means having an input encoding rate which could remain strictly positive or arbitrarily close to a capacity when the length of the code tends to infinity.

In a strongly constrained context (independent encoding of

operands for bit-by-bit boolean operations), P. Elias obtained negative first results on the existence of a noisy computation capacity ([6]). This work was deepened by Peterson and Rabin in [7] and by Winograd in [8]. One of the major conclusions of these studies was that reliable computation with positive rate (the ratio  $\frac{k}{n}$  of encoding  $k$ -length input block in  $n$ -length blocks of binary symbols) in the presence of noise is not possible for some boolean operations (e.g., AND) under some assumptions (independent coding of operands, bijective decoding and bit-by-bit operation). This led to the conclusion that, under these assumptions, there is no capacity for such noisy operations. It is worth noting that the assumptions were made to forbid the reliable encoder and decoder to "participate" to the computation of the expected operation.

In [9], Ahlswede went into the subject in greater depth and made an important contribution in characterizing contexts in which a capacity for noisy computations cannot exist. It appears that the characteristics of the decoding function play a fundamental role. If the inverse of the decoding function is injective and monotonic then weak converse theorems hold for the average and maximal error probabilities. If, in addition, the inverse of the decoding function preserves the logical AND (this implies monotonicity), then strong converses hold. These theorems state that the rate of encoding tends to 0 when the block code length tends to infinity. The hypotheses made in [6]–[8], i.e., independent encoding of operands and bijective decoding, imply monotonicity of the inverse of the decoding function. On this aspect, [9] supersedes [6]–[8].

Nevertheless, these negative results do not imply the absolute impossibility to identify a capacity for noisy computation. They characterize codes, encoding and decoding processes which cannot open this ability. To define a capacity for noisy computations, assumptions must be relaxed.

To the author's best knowledge, the first positive answer given through a definition of a capacity of a noisy computation (in fact similar to the one for a noisy channel) and a coding theorem came from Winograd and Cowan in [10]. In [10], the entropy  $H(X|F(X))$  of the input source conditioned by the noisy computation output is assessed as a noise measure. As it is the equivocation between the *noisy output* and the *input*, this quantity is not relevant, in full generality, as the equivocation due to the sole noise: it encompasses also the amount of information lost by computation. But, in a special case of noisy functions called decomposable modules,  $H(X|F(X))$  actually measures the equivocation due to noise. Decomposable modules are noisy functions which can be modeled by a

perfect function followed by a noisy communication channel: the error probability depends on the desired output value rather than on the input value. These peculiar functions, though noisy, make the context equivalent to that where the reliable encoder computes and encodes the expected function result before communication through a noisy channel. Due to the restriction of considering decomposable modules, [10] did not completely succeed in proposing a noisy computation capacity in a general scope ([10], theorem 6.3, pages 47-48).

Noisy computation capacity is also considered in reliable reconstruction of a function of sources over a multiple access channel. Much more recently, a definition of noisy computation capacity is established by Nazer and Gastpar in [11] and is totally consistent with the one proposed here. Nazer and Gastpar demonstrate the possible advantages of joint source-channel coding of multiple sources over a separation-based scheme, allowing a decoder to retrieve a value which is a function of input sources. This context makes relevant the proposed distributed encoding process which perfectly performs a computation equivalent to the desired function. The encoder outputs are then transmitted through a noisy MAC to a decoder (see proofs of Theorems 1 and 2 of [11]). This also models a noisy computation as a perfect computation followed by a noisy transmission of the result. It can be noticed that [10] and [11] relax the assumptions of [6]–[9] in a similar way: all goes as if the operands are jointly coded into an encoded form of the expected function result before being handled by a noisy communication channel.

The present paper establishes a model setting down the problem of noisy computation (section II), a definition of the capacity of a noisy computation with respect to an expected function and a coding lemma (section III). A model for reliable computation is given, section IV. Based on this model, a joint source-computation coding theorem and its converse are stated and proved in Section V. This theorem aims at formally capturing practical approaches in which reliable computation of a function  $g$  is obtained thanks to a noisy apparatus  $F$  computing with noise a function  $f$  (e.g., a regular arithmetic addition  $g$  obtained from the noisy actual circuit  $F$  implementing  $f$  which is an addition acting on residue encoded operands, [12]). The input source,  $f$ ,  $g$ ,  $F$  and the cascade  $f^{-1}F$  are supposed to be AMS and ergodic one-sided random processes or channels, extending [13] to more general random processes and algorithms. The perfect function  $f$  is assumed *unary* (as is a Turing computable function).  $n$ -ary functions can be modeled as unary ones by concatenating  $n$  input values in one "meta"-input and thus modeling a *joint coding* of operands. This relaxes the assumptions of [6]–[9].

## II. MODEL FOR NOISY COMPUTATION

In this section, the notations used follow [14] and cover countable alphabets, assumed standard (thus conditional probabilities are regular).

Let  $X \equiv \{X_i; i \in \mathcal{I}\}$  a random process with values in  $(A^{\mathcal{I}}, \mathcal{B}_{A^{\mathcal{I}}})$  where  $A$  is a countable alphabet and  $\mathcal{I}$  a countable set of indexes (e.g.,  $\mathbb{N}$ ).  $\mathcal{B}_{A^{\mathcal{I}}}$  denotes the  $\sigma$ -field generated by

the rectangles (chapter 1, [14]). The random process  $X$  is the source of inputs of the noisy computation.

The noisy computing device is modeled as a random channel, i.e. a set of conditional probabilities  $F \equiv \{F_x, x \in A^{\mathcal{I}}\}$ , taking  $X$  as input and producing as an output a random process  $Z \equiv \{Z_i; i \in \mathcal{I}\}$  on  $(C^{\mathcal{I}}, \mathcal{B}_{C^{\mathcal{I}}})$  where  $C$  is a countable alphabet. The hookup  $P_{XZ} \equiv P_X F$  is the probability measure characterizing the actual noisy computation with an input flow represented by  $X$ . From [14], chapter 2,  $\forall O \in \mathcal{B}_{A^{\mathcal{I}} \times C^{\mathcal{I}}}$ :

$$P_{XZ}(O) = \int_{A^{\mathcal{I}}} P_{Z|X}(O_x|x) dP_X = \int_{C^{\mathcal{I}}} P_{X|Z}(O_z|z) dP_Z$$

where  $O_x = \{z \in C^{\mathcal{I}} / (x, z) \in O\}$ . The probabilities  $\{P_{Z|X}(\cdot|x), x \in A^{\mathcal{I}}\}$  defines the channel  $X \rightarrow Z$  ( $F_x \equiv P_{Z|X}(\cdot|x)$ ) and  $\{P_{X|Z}(\cdot|z), z \in C^{\mathcal{I}}\}$  the "reverse" channel  $F^{-1}$ .  $P_{XZ}$  fully determines the channels  $F$  and  $F^{-1}$ . Conversely, if  $X$  and a set of conditional probabilities  $\{P_{Z|X}(\cdot|x), x \in A^{\mathcal{I}}\}$  (i.e.  $F$ ) are given, then  $P_{XZ}$  and the output process  $Z$  are well defined. A functional notation  $Z = F(X)$  will be used below.

The desired (i.e. perfect) computation will be represented by a *measurable* function  $f : A^{\mathcal{I}} \rightarrow B^{\mathcal{I}}$  where  $B$  is a countable alphabet.  $Y = f(X)$  is a random process of distribution  $P_Y = P_X f^{-1}$ . The function  $f$  defines a deterministic channel  $X \rightarrow Y = f(X)$  which is a set of conditional probabilities  $\{P_{f(X)|X}(\cdot|x), x \in A^{\mathcal{I}}\}$  (See [14], chap. 2):

$$\forall G \in \mathcal{B}_{B^{\mathcal{I}}}, P_{f(X)|X}(G|x) = 1_{f^{-1}(G)}(x) P_X \text{ a.e.} \quad (1)$$

$F$  and  $f$  determine a channel  $f(X) \rightarrow Z$  which is a cascade of the reverse channel  $f^{-1}$  followed by  $F$  (figure (1)). The noisy computation model should not be understood as

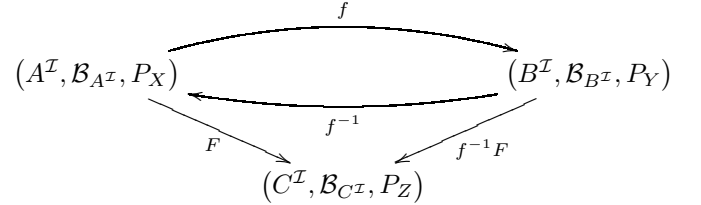


Fig. 1. Model for Noisy Computation

a cascade made of a perfect function followed by a noisy channel (as done in [10] for instance). The cascade  $f^{-1}F$  is an "artifact" on which a channel coding theorem will be invoked to build a code for the noisy computation.

From the expression giving the probabilities of a cascade of channels ([14], chap. 2) and applying (1), it easily comes that  $P_{Y|Z} \equiv P_{X|Z} f^{-1}$ .

If  $f$  is bijective and  $A = B$ , the model is the noisy channel one.

## III. CAPACITY OF A NOISY COMPUTATION

The alphabets are now assumed finite and the  $\sigma$ -fields  $\mathcal{B}_A$ ,  $\mathcal{B}_B$  and  $\mathcal{B}_C$  are the sets of subsets of  $A$ ,  $B$  and  $C$ .  $f^n(X^n)$  stands for the random value  $Y^n$  (the  $n$  first symbols of  $Y$ ),  $F^n(X^n)$  for  $Z^n$ .  $[A \times B, Xf; A \times C, XF]$  denotes the noisy

computation  $F$  of  $f$  on  $X$ .  $X$ ,  $F$ ,  $f$  and the cascade  $f^{-1}F$  are assumed AMS and ergodic. Thus entropy rates are limits and ergodic theorems hold ([14]).

The Feinstein's theorem is reminded ([14], chap. 12):

**Theorem 1 (Feinstein's theorem):** Let  $[A \times B, \mu\nu]$  be an AMS and ergodic hookup of a source  $\mu$  and channel  $\nu$ . Let  $\bar{I}_{\mu\nu} = \bar{I}_{\mu\nu}(X; Y)$  denote the average mutual information rate and assume that  $\bar{I}_{\mu\nu}$  is finite (this is the case if the alphabets are finite.). Then for any  $R < \bar{I}_{\mu\nu}$  and for any  $\epsilon > 0$ , there exists, for  $n$  large enough, a code  $\{(\omega_i, \Gamma_i) \in A^n \times \mathcal{B}_{B^n}, i = 1, \dots, M\}$  such that  $M = \lfloor e^{nR} \rfloor$  and  $\forall i = 1, \dots, M, \hat{\nu}^n(\Gamma_i^c | \omega_i) \leq \epsilon$  and  $\hat{\nu}^n$  is the channel induced by the source  $\mu$ , ([14], chap. 12).

The following definition introduces the *typical input rate* which will be shown to be the rate at which a source should produce typical inputs for a random process to allow to recover the desired function results by decoding.

**Definition 1:** The *typical input rate* of the source  $X$  for the noisy computation  $F$  with respect to the perfect function  $f$  is the following limit, denoted  $\bar{B}(X, f, F)$ :

$$\lim_{n \rightarrow \infty} \frac{H(X^n | f^n(X^n)) + I(F^n(X^n); f^n(X^n))}{n}$$

Taking into account the hypothesis (AMS and ergodic source and channels),  $\bar{B}(X, f, F)$  is well defined. Simple algebra gives  $\bar{B}(X, f, F) = \bar{H}(X) - \bar{H}(f(X) | F(X))$

**Definition 2:** Let  $[A \times B, Xf; A \times C, XF]$  be a noisy computation on finite alphabets  $A$ ,  $B$  and  $C$ . A  $[M, n, \epsilon]$ -Feinstein code for the noisy computation  $[A \times B, Xf; A \times C, XF]$  is a set  $\{(A_i^n, \Gamma_i^n) \in \mathcal{B}_{A^n} \times \mathcal{B}_{C^n}, i = 1, \dots, M\}$  such that:

- 1)  $\hat{P}_{F(X)|X}^n(\Gamma_i^n | x^n) \leq \epsilon$  for any  $x^n \in A_i^n, i = 1, \dots, M$
- 2)  $\forall i = 1, \dots, M, \exists y_i \in B^n$  such that  $A_i^n = (f^n)^{-1}(y_i)$

$F^n(X^n)$   $\epsilon$ -reliably computes  $f^n(X^n)$  on the code  $\{(A_i^n, \Gamma_i^n), i = 1, \dots, M\}$ .

**Lemma 1 (Feinstein's theorem for Noisy Computation):**

Let  $[A \times B, Xf; A \times C, XF]$  be a noisy computation on finite alphabets. For any  $R < \bar{B}(X, f, F)$ , for any  $\epsilon > 0$ ,  $n$  large enough, there exists a  $[\lfloor e^{n(R - \bar{H}(X|f(X))} \rfloor, n, \epsilon]$ -Feinstein code for  $F^n(X^n)$  to  $\epsilon$ -reliably computes  $f^n(X^n)$ .

**Proof:**  $R < \bar{B}(X, f, F) \Rightarrow R' = R - \bar{H}(X|f(X)) < \bar{I}(f(X), F(X))$ . Then, thanks to the Feinstein's theorem, since  $P_{f(X), F(X)}$  is AMS and ergodic (by assumption), for  $n$  large enough, there exists a Feinstein code  $\{(y_i, \Gamma_i) \in B^n \times \mathcal{B}_{C^n}; i = 1, \dots, M\}$  such that  $M = \lfloor e^{nR'} \rfloor$  and  $\forall i = 1, \dots, M; \hat{P}_{F(X)|X}^n(\Gamma_i^c | y_i) \leq \epsilon$

Let  $x$  belong to  $(f^n)^{-1}(y)$ . Considering the cascade  $X^n \rightarrow f^n(X^n) \rightarrow F^n(X^n)$ , for any  $k = 1, \dots, M$ :

$$\hat{P}_{F(X)|X}^n(\Gamma_k^c | x) = \int_{B^X} \hat{P}_{F(X)|f(X)}^n(\Gamma_k^c | \underline{y}) d\hat{P}_{f(X)|X}^n(\underline{y} | x)$$

$$\begin{aligned} \hat{P}_{F(X)|X}^n(\Gamma_k^c | x) &= \int_{\{y_k\}} \hat{P}_{F(X)|f(X)}^n(\Gamma_k^c | \underline{y}) d\hat{P}_{f(X)|X}^n(\underline{y} | x) \\ &+ \int_{\{y_k\}^c} \hat{P}_{F(X)|f(X)}^n(\Gamma_k^c | \underline{y}) d\hat{P}_{f(X)|X}^n(\underline{y} | x) \\ &\leq \hat{P}_{f(X)|X}^n(\{y_k\} | x) \cdot \epsilon + \hat{P}_{f(X)|X}^n(\{y_k\}^c | x) \end{aligned}$$

If  $x \in (f^n)^{-1}(y_k)$  then  $\hat{P}_{f(X)|X}^n(\{y_k\} | x) = 1$  and  $\hat{P}_{f(X)|X}^n(\{y_k\}^c | x) = 0$ , hence  $\forall x \in (f^n)^{-1}(y_k)$ ,  $\hat{P}_{F(X)|X}^n(\Gamma_k^c | x_k) \leq \epsilon$  ■

We can conclude this section by the definition of the *typical input capacity* of a noisy computation.

**Definition 3:** The typical input capacity of the noisy function  $F$  with respect to the perfect function  $f$  is  $C_f(F) = \sup_{\text{AMS erg } P_X} \bar{B}(X, f, F)$ , the supremum is over all AMS and ergodic sources  $X$ .

The equivalent expression  $C_f(F) = \sup_{\text{AMS erg } P_X} [\bar{H}(X) - \bar{H}(f(X) | F(X))]$  shows that this capacity boils down to the "usual" channel capacity when  $f$  is a bijection, in which case  $\bar{H}(f(X) | F(X)) = \bar{H}(X | F(X))$ .

#### IV. RELIABLE COMPUTATION

There is a need ([6], [8]) to constrain the encoding and decoding processes to avoid the following cases:

- either an (assumed perfect) encoder which computes the expected function, encodes the result before transmission through the random process (considered as a noisy transmission channel)
- or an encoder which encodes input values for reliable transmission through the random process (considered here also as a noisy transmission channel) and a decoder (assumed also reliable) which decodes (almost perfectly) and computes (perfectly) the expected function.

Considering that a computation  $g$  is a "true" computation if the entropy is reduced ( $\bar{H}(X' | g(X')) > 0 \Rightarrow \bar{H}(g(X')) < \bar{H}(X')$  (else it is communication), the model must be targeted to be mainly relevant for non-injective functions (i.e.  $\bar{H}(X' | g(X')) > 0$ ). For injective  $g$ , this becomes the classical reliable transmission model. With the constraint that both encoding and decoding are based on injections (in a sense made precise below) then the encoder and the decoder cannot compute (at least totally) the desired function as they do not reduce entropy.

The proposed model of the complete process to reliably compute a function  $g : A'^{\mathcal{I}} \rightarrow B'^{\mathcal{I}}$  acting on a source  $X'$ , thanks to a noisy implementation  $F$  of a function  $f : A^{\mathcal{I}} \rightarrow C^{\mathcal{I}}$  is the following:

- **encoding:** let  $X^n$  be the  $n^{\text{th}}$  extension of a source for which we have a maximal code  $(A_i^n, \Gamma_i)_{i=1, \dots, M}$  allowing to  $\epsilon$ -reliably compute  $f^n(X^n)$  by  $F^n(X^n)$  (cf lemma 1 and definition 2) ; a typical  $k$ -sequence  $x'$  of  $X'^k$  is encoded into a *typical given*  $y_i$  (this important assumption is discussed in the conclusion)  $n$ -sequence of  $X^n$  by an injective function, say  $\mathcal{U}$ , such that  $\mathcal{U}(x') \in A_i^n$  for some  $i = 1, \dots, M$
- **computation of the noisy function:**  $F^n$  is applied to  $\mathcal{U}(x')$  producing a typical  $n$ -sequence  $F^n(\mathcal{U}(x'))$  of  $F^n(X^n)$  where  $F^n(\mathcal{U}(x'))$  belongs to a given  $\Gamma_i$  (with probability greater than  $1 - \epsilon$ )
- **decoding:** the first step is to associate to  $F^n(\mathcal{U}(x'))$  the typical  $n$ -sequence  $y_i$  of  $f^n(X^n)$  corresponding to  $\Gamma_i$ , the second step is to apply to  $y_i$  a function

$\mathcal{V} : \{\mathbf{y}_1, \dots, \mathbf{y}_M\} \rightarrow \{\text{typical } k\text{-sequences of } g^k(X'^k)\}$   
such that  $\mathcal{V}(y_i) = g^k(x')$

A decoding error occurs when one obtains a  $n$ -sequence  $y_j$  (or equivalently a  $\Gamma_j$ ) such that  $\widehat{g^k(x')} = \mathcal{V}(y_j) \neq g^k(x')$

To be able to define a decoding function  $\mathcal{V}$  (i.e., a deterministic decoding), the encoding function  $\mathcal{U}$  has to be such that the typical (given  $y_i$ )  $n$ -sequences of one  $A_i^n = (f^n)^{-1}(y_i)$  ( $y_i \in \{y_1, \dots, y_M\}$ ) are used for encoding typical  $k$ -sequences of *only one*  $(g^k)^{-1}(z)$ ,  $z$  typical  $k$ -sequence of  $g^k(X'^k)$ .

We also require that  $\mathcal{V}$  be an injection (as we have required from  $\mathcal{U}$ ).

The typical  $k$ -sequences of a  $(g^k)^{-1}(z)$ ,  $z$  typical  $k$ -sequence of  $g^k(X'^k)$ , are encoded in typical (given  $y_i$ )  $n$ -sequences of one and only one  $A_i^n$ . So, if  $x'_1$  and  $x'_2$  are two typical  $k$ -sequences of  $X'^k$ :

$$f^n(\mathcal{U}(x'_1)) = f^n(\mathcal{U}(x'_2)) \Leftrightarrow g^k(x'_1) = g^k(x'_2)$$

The model fulfills the constraints identified above. The encoder implements an injection and thus cannot compute the desired function  $f$  nor  $g$  (if  $f$  and  $g$  are not injective). The same comment applies to the injective decoding step  $\mathcal{V}$ .

## V. A CODING THEOREM AND ITS CONVERSE

The sources, functions, noisy function and the cascade  $f^{-1}F$  are assumed AMS and ergodic.

*Definition 4:* With the notations of section IV, the ratio  $R = \frac{k \cdot H(X')}{n}$  is called the *typical encoding input rate*. A rate  $R$  is said to be *achievable with respect to the function  $f$*  if there exists a sequence of codes of size  $n$  such that the maximal probability of decoding error tends to 0 as  $n$  tends to infinity.

*Theorem 2:* If  $R < C_f(F)$ , then  $R$  is achievable w.r.t  $f$ .

*Proof:*

This proof, although identical to that in [13], is given as it includes the starting point for the proof of the converse theorem. First, it is shown that the injective encoding of typical  $k$ -sequences of a set  $(g^k)^{-1}(z)$  on typical (given  $y_i$ )  $n$ -sequences belonging to  $A_i^n$  is possible for suitably chosen  $k$  and  $n$  (lossless coding). Secondly, it is shown that, at encoding input rates below capacity and for  $k$  and  $n$  suitably chosen, the sets  $A_i^n$  are almost as many as the sets  $(g^k)^{-1}(z)$ .

Let  $\delta'' > 0$ . Since  $\mathbb{Q}$  is dense in  $\mathbb{R}$ , there exist  $k$  and  $n$  such that:

$$\frac{\overline{H}(X'|g(X'))}{\overline{H}(X|f(X))} < \frac{n}{k} < \frac{\overline{H}(X'|g(X')) + \delta''}{\overline{H}(X|f(X))}$$

Moreover,  $k$  and  $n$  can be chosen as large as needed. Thus:

$$\frac{k \cdot \overline{H}(X'|g(X'))}{H(X|f(X))} < n < \frac{k \cdot (\overline{H}(X'|g(X')) + \delta'')}{\overline{H}(X|f(X))} \quad (2)$$

We can choose  $\delta, \delta' > 0$  and  $0 < \epsilon < 1/2$  small enough for:

$$\begin{aligned} \frac{k \cdot (\overline{H}(X'|g(X')) + \delta) - \log(1 - 2\epsilon)}{\overline{H}(X|f(X)) - \delta'} &< n \\ &< \frac{k \cdot (\overline{H}(X'|g(X')) + \delta + \delta'')}{\overline{H}(X|f(X)) + \delta'} \end{aligned}$$

giving

$$\begin{aligned} k \cdot (\overline{H}(X'|g(X')) + \delta) &< \log(1 - 2\epsilon) + n \cdot (\overline{H}(X|f(X)) - \delta') \\ &< n \cdot (\overline{H}(X|f(X)) + \delta') < k \cdot (\overline{H}(X'|g(X')) + \delta + \delta'') \end{aligned}$$

If  $\nu_1$  is the number of typical  $k$ -sequences of  $(g^k)^{-1}(z)$  and  $\nu_2$  is the number of typical (given  $y_i$ )  $n$ -sequences in an  $A_i^n$ , we have (by conditional AEP):

$$\begin{aligned} \nu_1 &< e^{k \cdot (\overline{H}(X'|g(X')) + \delta)} < (1 - 2\epsilon) e^{n \cdot (\overline{H}(X|f(X)) - \delta')} \\ &< \nu_2 < e^{n \cdot (\overline{H}(X|f(X)) + \delta')} < e^{k \cdot (\overline{H}(X'|g(X')) + \delta + \delta'')} \end{aligned}$$

It is thus possible to find an injection from the set of typical  $k$ -sequences of  $(g^k)^{-1}(z)$  on the subset of typical sequences (given  $y_i$ ) of  $A_i^n$ . This shows the first step.

Assume that  $R = k \overline{H}(X')/n < \overline{H}(X) - \overline{H}(f(X)|F(X)) \leq C_f(F)$ . Such a  $X$  exists by definition of  $C_f(F)$ . So

$$\begin{aligned} k(\overline{H}(g(X')) + \overline{H}(X'|g(X'))) &< \\ n \cdot (\overline{H}(f(X)) - \overline{H}(f(X)|F(X))) + n \cdot \overline{H}(X|f(X)) \end{aligned}$$

By (2),  $n \cdot \overline{H}(X|f(X)) - k \cdot \overline{H}(X'|g(X')) < k \cdot \delta''$  thus

$$k \overline{H}(g(X')) < n \cdot (\overline{H}(f(X)) - \overline{H}(f(X)|F(X))) + k \cdot \delta''$$

$\epsilon_1, \delta''' > 0$  can be chosen small enough in order to get:

$$e^{k(\overline{H}(g(X')) + \delta''')} < e^{n \cdot (\overline{H}(f(X)) - \overline{H}(f(X)|F(X)) + \frac{k}{n} \cdot \delta'' - \epsilon_1)}$$

If  $\nu_3$  is the number of typical  $k$ -sequences of  $g^k(X'^k)$  and  $M$  is the size of the code (i.e., the number of  $(A_i^n, \Gamma_i)$ ), we have (by AEP and Lemma 1):

$$\begin{aligned} \nu_3 &< e^{k(\overline{H}(g(X')) + \delta''')} \\ &< e^{n \cdot (\overline{H}(f(X)) - \overline{H}(f(X)|F(X)) + \frac{k}{n} \cdot \delta'' - \epsilon_1)} < M \end{aligned}$$

The assumed model, by the constraints on encoding, implies that the best ratio (i.e., the smaller)  $\frac{n}{k}$  of encoding respects the inequality (2):  $\frac{\overline{H}(X'|g(X'))}{\overline{H}(X|f(X))} < \frac{n}{k} < \frac{\overline{H}(X'|g(X')) + \delta''}{\overline{H}(X|f(X))}$ . Let  $\gamma = \frac{\overline{H}(X'|g(X'))}{\overline{H}(X|f(X))}$ . To respect the encoding constraints (typical sequences are "injectively" encoded into typical sequences), a rate  $R = \frac{k}{n} \overline{H}(X')$  must be such that  $R \leq \frac{\overline{H}(X')}{\gamma}$

*Theorem 3:* If  $R > C_f(F)$ , there is no code such that the error probability tends to 0 as  $n \rightarrow \infty$

*Proof:* The decoding is deterministic then:

- 1)  $F^n(X^n) \rightarrow f^n(X^n) \rightarrow g^k(X'^k)$  is a Markov Chain thus  $g^k(X'^k) \rightarrow f^n(X^n) \rightarrow F^n(X^n)$  is a Markov Chain
- 2)  $f^n(X^n) \rightarrow F^n(X^n) \rightarrow \widehat{g^k(X'^k)}$  is a Markov Chain

Hence  $g^k(X'^k) \rightarrow f^n(X^n) \rightarrow F^n(X^n) \rightarrow \widehat{g^k(X'^k)}$  is a Markov Chain. This implies that,  $\forall n, k$  such that  $\frac{k}{n} \leq \frac{1}{\gamma}$ :

$$\begin{aligned} I(g^k(X'^k); \widehat{g^k(X'^k)}) &\leq I(f^n(X^n); F^n(X^n)) \text{ hence} \\ H(g^k(X'^k)) - H(g^k(X'^k)|\widehat{g^k(X'^k)}) &\leq \\ I(f^n(X^n); F^n(X^n)) \\ \Rightarrow H(X'^k) - H(X'^k|g^k(X'^k)) - H(g^k(X'^k)|\widehat{g^k(X'^k)}) &\leq \\ I(f^n(X^n); F^n(X^n)) \end{aligned}$$

by Fano's inequality:

$$\begin{aligned} H(X'^k) - H(X'^k|g^k(X'^k)) &- (H_2(P_e(k)) + k \cdot P_e(k) \cdot \log(|B'|)) \\ &\leq I(f^n(X^n); F^n(X^n)) \\ \Rightarrow \frac{H(X'^k)}{k} - \frac{H_2(P_e(k))}{k} - P_e(k) \cdot \log(|B'|) &\leq \\ \frac{H(X'^k|g^k(X'^k))}{k} + \frac{I(f^n(X^n); F^n(X^n))}{k} \\ \Rightarrow \frac{H(X'^k)}{k} - \frac{H_2(P_e(k))}{k} - P_e(k) \cdot \log(|B'|) &\leq \\ \frac{H(X'^k|g^k(X'^k))}{k} + \gamma \cdot \frac{I(f^n(X^n); F^n(X^n))}{n} \end{aligned}$$

If the error probability is asymptotically 0 (i.e.,  $\lim_{k \rightarrow \infty} (P_e(k) = 0)$ ) then necessarily (letting  $k$  and  $n$  tend to infinity):  $\overline{H}(X') \leq \overline{H}(X'|g(X')) + \gamma \cdot \overline{I}(f(X); F(X))$ . But  $\overline{H}(X'|g(X')) = \gamma \overline{H}(X|f(X))$ , then:

$$\overline{H}(X') \leq \gamma \cdot (\overline{H}(X|f(X)) + \overline{I}(f(X); F(X)))$$

since  $R \leq \frac{\overline{H}(X')}{\gamma}$ , we obtain

$$R \leq \overline{H}(X|f(X)) + \overline{I}(f(X); F(X)) \leq C_f(F)$$

Thus if  $R > C_f(F)$  then the error probability does not vanish. ■

## VI. DISCUSSION AND CONCLUSION

The coding lemma, the coding theorem and its converse assume that the sources ( $X$  and  $X'$ ), the channels ( $F$ ,  $f$ ,  $g$ ) and the cascade  $f^{-1}F$  are AMS and ergodic. Cases can be identified where it is possible to derive the AMS property and ergodicity of  $f^{-1}F$  from properties of  $X$ ,  $F$  and  $f$  (e.g., if  $X$ ,  $F$  are stationary and weakly mixing and  $f$  AMS and ergodic then  $f^{-1}F$  is AMS and ergodic). Due to lack of space, this question is not addressed here, neither the identification of classes of AMS and ergodic functions  $f$  and  $g$ .

The model of reliable computation assumes that the encoder and the decoder are perfectly reliable. This assumption could be justified by quoting from [8] "*The computation system [model] was devised for the sole purpose of studying the relation of information theory of reliable automata*". Moreover we could argue in addition that if the complexity of the computation device is of a much greater magnitude than that of the encoder and decoder then the unreliability of the encoder and decoder have almost no impact on the overall reliability

of the computation and thus can be neglected. For complex systems (e.g., based on significant software volume), this is quite realistic. In any case, it is impossible to overcome the fact that the reliability reached is at the best the reliability of the final decoding device. The only way is to built a intrinsically reliable enough decoder (for example thanks to gate redundancy). A noisy encoder is a noisy computation itself and thus can be handled from the point of view of "cascaded noisy computations". This is outside of the scope of the present paper.

The proposed model of reliable computation involves two perfect functions  $g$  and  $f$ . This is intended to capture major real cases as already mentioned. Another motivation to use an "ancillary" function  $f$  in the model is that this is an efficient way to define an input code, meaning a family of subsets  $(A_i^n)_{i=1, \dots, M}$ , that do not overlap and whose "images" by the noisy function do not overlap "too much" (i.e., fall into disjoint  $\Gamma_i$  with high probability). Defining such family is defining (partially) a function  $f$  by picking, for each  $i$  an  $y_i$  and stating  $f^{-1}(y_i) = A_i$ . In addition  $f$  allows a characterization of a kind of size of the sets  $(f^n)^{-1}(y_i)$  through the conditional entropy rate  $\overline{H}(X|f(X))$ . This motivates also the constraint of coding by conditionally typical sequences. While the sets  $(f^n)^{-1}(y_i)$  are "balanced", for large  $n$ , with respect to the number of (conditionally) typical sequences they contain, their cardinalities might be very different and bounds are not straightforward to obtain. Thus, the use of all possible elements of  $(f^n)^{-1}(y_i)$  forbids to characterize all the  $(f^n)^{-1}(y_i)$  by the same number measuring the "encoding" power. The same difficulty forbids to state a converse as well. The "encoding by conditionally typical sequences" trick overcomes this difficulty.

## REFERENCES

- [1] D. A. Spielman, "Highly fault-tolerant parallel computation," in *Annual Symposium Foundations of Computer Science*, 1996, pp. 154–160.
- [2] P. Gacs, "Reliable computation," Boston University, Tech. Rep., 2005.
- [3] C. N. Hadjicostis and G. C. Verghese, "Coding approaches to fault tolerance in linear dynamic systems," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 210–228, january 2005.
- [4] E. Rachlin and J. E. Savage, "A framework for coded computation," in *ISIT 2008*, 2008.
- [5] J. V. Neumann, "Probabilistic logics and the synthesis of reliable organisms from unreliable components," *Automata studies*, 1956.
- [6] P. Elias, "Computation in the presence of noise," *IBM Journal*, october 1958.
- [7] W. Peterson and M. Rabin, "On codes for checking logical operations," *IBM Journal*, april 1959.
- [8] S. Winograd, "Coding for logical operations," *IBM Journal*, october 1962.
- [9] R. Ahlswede, "Improvements of Winograd's result on computation in the presence of noise," *IEEE Transactions on information theory*, vol. IT-30, no. 6, november 1984.
- [10] S. Winograd and J. Cowan, *Reliable computation in presence of noise*. The MIT Press, 1963.
- [11] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Transactions on information theory*, vol. 53, no. 10, october 2007.
- [12] T. Rao and E. Fujiwara, *Error-control coding for computer systems*. Prentice-Hall, 1989.
- [13] F. Simon, "Capacity of a noisy function," in *Information Theory Workshop - Dublin*, september 2010.
- [14] R. M. Gray, *Entropy and Information Theory*, 2nd edition. Springer, 2011.